Supplementary Methods

Whole Exome Sequencing

Preparation of sequencing libraries Genomic DNA libraries were prepared using TruSeq DNA library kit (Illumina) followed by exome capture using SureSelect XT Human All Exon V5 (Agilent). Library preparation and exome capture were performed according to manufacturer protocol with modifications as described(4).

Library sequencing and analysis. Libraries were multiplexed onto Illumina HiSeq 2500 and paired-end sequenced with read lengths of 2 x 100 bp. Raw sequencing data was processed through Illumina secondary analysis software CASAVA 1.8 with ELANDv2 alignment to hg19. Sequence alterations were determined using stringent quality criteria comparing tumor and normal DNA. For each sample, neoplastic content was estimated based on the average of mutation allele fractions identified for each sample.

Confirmation of somatic mutations. All known cancer driver gene mutations were independently validated by Sanger sequencing of the tumor and matched normal tissue. Additionally, seven to nine mutations were randomly chosen for each sample to confirm by Sanger sequencing. Mutations that were tested by Sanger sequencing are noted in Supplementary Table S3.

Supplementary Figure 1. Copy number alterations in ten AA-exposed Taiwanese ccRCCs. Circos plots derived from allelic fractions (0 to 1 on y-axis)

of germline SNPs from whole-exome sequencing. Black outer circle represents banding pattern of chromosomes 1-22, X and Y (if male), middle red circle is normal (germline), and inner blue circle is tumor. In normal tissues SNPs should theoretically be heterozygous (allelic fraction of 0.5) or homozygous (allelic fraction of 1). However, tumor allelic fractions may differ from germline due to loss of heterozygosity, tumor heterogeneity, copy number variation, normal contamination, or noise. **Methods**. Germline single-base substitutions present in both normal and tumor DNA samples for each patient were identified from the whole exome sequencing data. Allelle fraction (AF), defined as number of distinct reads supporting the alternate allele as a proportion of the total of distinct reads, was calculated for each SBS in both the normal and tumor. AF for each substitution was represented graphically for each patient using CIRCOS(30). Each plot contains an ideogram of chromosome 1-22, X, and where appropriate, Y. Normal (germline) AF for each substitution is plotted in the outer red circle and tumor AF is plotted in the inner blue circle. In each case, the inner and outer most lines represent AF of 0 and 1, respectively. Each line between 0 and 1 represent an AF increment of 0.1.